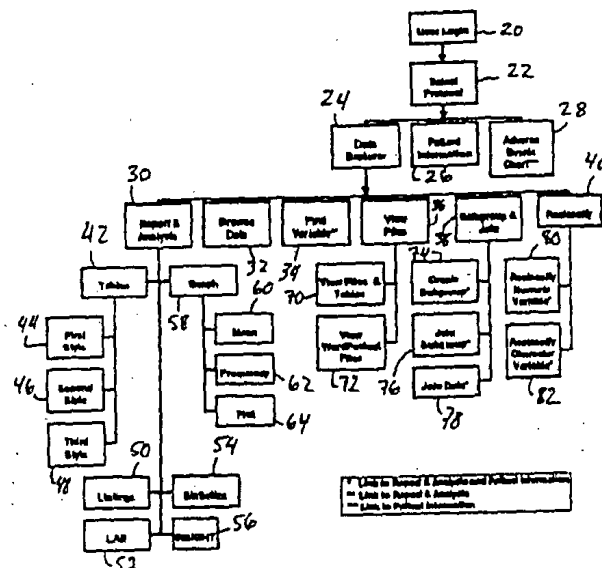




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>G06T 11/00, G06F 17/30</b>		A1	(11) International Publication Number: <b>WO 98/12669</b>
			(43) International Publication Date: 26 March 1998 (26.03.98)
(21) International Application Number: <b>PCT/US97/16629</b>		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 18 September 1997 (18.09.97)			
(30) Priority Data: 60/026,322 18 September 1996 (18.09.96) US		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(71) Applicant: PHARM-DATA, INC. [US/US]; Route 659 Oakhurst Farm Road, Kent's Store, VA 23084-0070 (US).			
(72) Inventors: KANTER, Donald, R.; 2034 Kinridge Trail, Marietta, GA 30062 (US). FINN, Andrew, L.; 3646 Sope Creek Farm, Marietta, GA 30067 (US). SAWYER, William, T.; 1006 Orange High Road, Hillsborough, NC 27278 (US). CHAO-YING, Hsieh; 821 Soaring Drive, Marietta, GA 30062 (US). HOUSER, Vincent, P.; 4436 Dobbs Crossing, Marietta, GA 30068 (US).			
(74) Agents: NEEDLE, William, H. et al.; Needle & Rosenberg, 127 Peachtree Street, N.E., Atlanta, GA 30303 (US).			

(54) Title: METHOD AND APPARATUS FOR PROCESSING CLINICAL TRIAL DATABASES



## (57) Abstract

A method and apparatus for facilitating review of clinical data provides a point-and-click, menu-driven approach for reviewing, analyzing, and graphing clinical data (30, 32, 34, 36, 38 and 40). Real-time access to clinical information in electronic databases is provided so as to permit clinicians to obtain information in a timely manner. A user is provided with the ability to browse patient profiles (32), perform data queries, reclassify numeric and character variables, create and analyze variable subgroups (74), create basic summary tables (42) and graphic data displays (58). The invention eliminates the time-consuming and problematic conversion step. The invention has the ability to export statistical databases to popular existing data formats. No programming experience is required for using this system.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BH	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NI	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CJ	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## METHOD AND APPARATUS FOR PROCESSING CLINICAL TRIAL DATABASES

5

### REFERENCE TO MATERIAL SUBJECT TO COPYRIGHT PROTECTION

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent disclosure, as it appears in the Patent and  
10 Trademark Office patent files or records, but otherwise reserves all copyrights whatsoever.

### REFERENCE TO A MICROFICHE APPENDIX

A microfiche appendix containing program code and a user's manual  
15 corresponding to the program code, comprising nine fiches and 580 frames, is submitted herewith.

## BACKGROUND OF THE INVENTION

### 20 1. Field of the Invention

The present invention relates to databases and, more specifically, to a method of processing clinical trial databases for users without a background in statistics.

### 25 2. Description of the Prior Art

Generally, examination of information in clinical trial databases has been accomplished either by manual tabulation of information contained in written reports summarizing the database, or by creating project-specific or inquiry-specific computer  
30 programs to perform selected functions to extract desired information. The manual cross-tabulation procedures are inadequate under normal circumstances, because the questions posed by the user are usually of such a specific nature that they are not likely to be answered by reorganization of information routinely contained in normal written

reports. At the same time, where users are required to create project-specific computer programs to extract information from these databases, the facility of the exploratory analysis may be hampered either by the user's lack of programming sophistication and experience, or by the need of the user to involve at least one other person (i.e., a programmer or biostatistician) in the process of creating necessary programming code. Even then, certain additional tasks are required. These tasks typically include: (1) writing computer code specific to questions to be asked of the database; (2) creating unique formats for tables and figures, to contain information summarized from the database; (3) preparing additional computer programming code to facilitate the exportation of summarized information (i.e., statistical output, tables, figures) to word processing or other application's software, for preparation of reports and other summaries of the information. Generally, this process may take several days to develop the programming code and create the database inquiry.

These methods have the disadvantages of inefficiency in the manual cross-tabulation process and loss of precision associated with the process of computer programming by a third party. In particular, the original question of interest to the user may not be adequately or completely answered by the computer programming code because of: (1) misunderstandings by the programmer of the user's desired information; (2) inability of the user to articulate to another individual exactly what they are interested in learning; (3) or incompatibilities between database software and word processing software. Furthermore, where users have an immediate need for information and data summarization, a process which requires several days for creation of computer code in order to answer questions of interest may significantly limit the user's ability to employ data and information efficiently and effectively in a clinical research decision-making.

In addition, where information contained in a clinical trial database is sensitive, or subject to restrictions on access limited to only a few individuals within an organization, the need to involve computer programmers and biostatisticians, who may not be authorized to have access to the database, creates problems with confidentiality

of information which may compromise the effective conduct of business by the organization using the database.

Other computer-based systems have been developed to provide the facility for 'real-time' interrogation of electronic databases. However, these systems are frequently designed by computer programmers and biostatisticians, and are intended for use by this same population of individuals, rather than by clinical personnel without substantive statistical or programming background. Consequently, the computer program interface between the database and the user may confuse or intimidate the user, who is neither a programmer nor a biostatistician.

Thus, there exists a need for a method and apparatus to facilitate efficient, reliable, and rapid end-user exploration of clinical trial databases to answer specific questions related to the information contained in the database. There also exists a need for a process in which the end-user directly interfaces with the database, using software that offers the ability to cross-tabulate information contained in the database, perform summary descriptive statistical analyses, and generate associated tables and charts.

### SUMMARY OF THE INVENTION

The disadvantages of the prior art are overcome by the present invention, which in one aspect is a method for exploring, examining, and summarizing information contained in an electronic database. Furthermore, because the invention provides an interface between the non-statistician, non-programmer user and the database, the invention provides for real time examination of databases. The invention provides the user with the ability to create subgroups and subsets of the data, merge these identified subgroups, reclassify data contained in the database, and create summary reports, including tables and graphs, of the data.

The invention includes a window-driven application designed for clinical data review. This system is highly intuitive and user-friendly and provides a point-and-click

menu-driven approach for reviewing, analyzing, and graphing clinical data. Potential users include FDA clinical reviewers of computer aided new drug applications and clinical staff at pharmaceutical companies. The invention permits clinicians to obtain information in a timely manner.

5

The invention is designed to accommodate different database structures. It automatically recognizes character and numeric variables to create different inquiry statements, which are important for programmers and biostatisticians, but not for users. The invention shows formatted values and labels, instead of raw values, which have limited meaning to users. It allows users to change variable and data set names into more descriptive names and to label the data sets and variables. It validates the database structure to be used in different modules to avoid user mistakes. It accommodates different types and names of key variables automatically for data merging once they are set up by an administrator.

15

An administrator controls the information to be reviewed, including study protocol, drug name, and indication levels; establishes user identification, password, and working directories for each user; arranges data set and variable names to be used in the Adverse Event Module; and arranges a key variable for data merging and subgrouping. This setup ensures system security and integrity.

20

The user can perform complicated inquiries without writing any code with the SAS® programming language. Traditionally, the user needed to know SAS programming, data format conventions, and database terminology in order to subset a data set in SAS. With the Subgrouping Module, the user can extract data with criteria he sets up online. The user can also create a subgroup with only the subject identification list and can use the data joining function later.

25

Using the SAS/GRAPH® Module in SAS is considered time-consuming and cumbersome, even for the most experienced SAS programmers. The Graph Module of the invention provides a way to create simple but informative graphs, which allows the

30

user to graphically present data trends, to drill down for detail data listings or spot information, and to export the graphs into a lot of popular graphic formats such as .bmp, .gif, .pcx, etc.

5           The Table Module in the invention can be used to generate three types of summary tables. It provides descriptive statistics such as means, standard deviations, and number of observations. The third style table module provides the option to count the number of patients or the number of observations in the output table.

10           After an administrator sets up an adverse event file and variable names, the Adverse Event Module allows the user to view adverse events by unique patient count instead of observation count. It also separates treatments, body systems, and preferred terms, and provides percent information by treatment for comparison between different treatments. Features like these that usually require extensive data manipulation and table  
15           programming are simple when using the invention.

          The Reclassify Module provides a way of viewing the data from a different perspective. It translates data into a new grouping convention while maintaining format type. It automatically provides data range information such as maximum and minimum  
20           for numeric data and data values information for character data.

          Most of the modules mentioned above are linked to the primary data browsing and analysis modules to provide more flexibility and power. Also, data set and variable labels are displayed throughout the explorer to provide more information.

25           Thus, it is the object of this invention to provide a method and apparatus for real time examination and summarization of electronic databases. In particular, it is the object of this invention to provide a mechanism by which non-statisticians, non-programmer users can directly access information in electronic databases, without the  
30           requisite need to create computer programs for this purpose. Thus time wasted because of the necessity of involving a computer programmer to generate code for exploration of

databases is avoided. Similarly, because the end user directly accesses the database, using an intuitive computer interface, the invention does not require significant training of end users in biostatistics or computer programming.

5           These and other aspects of the invention will become apparent from the following description of the preferred embodiments taken in conjunction with the following drawings. As would be obvious to one skilled in the art, many variations and modifications of the invention may be effected without departing from the spirit and scope of the novel concepts of the disclosure.

10

#### BRIEF DESCRIPTION OF THE FIGURES OF THE DRAWINGS

FIG. 1 is a flow chart showing the organization of the user-accessible modules of one embodiment of the invention.

15

FIG. 2 is a block diagram of a hardware configuration upon which a disclosed embodiment of the invention may run.

#### DETAILED DESCRIPTION OF THE INVENTION

20

A preferred embodiment of the invention is now described in detail. As used in the description herein and throughout the claims, the following terms take the meanings explicitly associated herein, unless the context clearly dictates otherwise: "a," "an," and "the" includes plural reference, "in" includes "in" and "on."

25

The invention may be embodied in a software program running in a digital computer. The complete source code for this embodiment is disclosed in the microfiche appendix, along with a user's guide that instructs the user how to operate all of the features of the embodiment.

30



As shown in FIG. 1, the invention includes a series of program modules 10, or subroutines, each of which performs specific functions. These include a user login 20, and a select protocol module 22. The user is allowed to select between a data explorer module 24, a patient information module 26 and an adverse events charting module 28.

5 The data explorer module 24 allows the user to select between a report and analysis module 30, a browse data module 32, a find variable module 34, a view files module 36, a subgroup and join module 38, and a reclassify module 40.

The report and analysis module 30 allows the user to select between a tables  
10 module 42, a listings module 50, a lab module 52, a statistics module 54, an INSIGHT module 56 and a graph module 58. The tables module 42 allows the user to select tables to be generated in one of three styles: a first style 44, a second style 46, and a third style 48. The graph module 58 allows the user to select between a mean module 60, a frequency module 62 and a plot module 64.

15

The view files module 36 allows the user to select between a view files & tables module 70 and a view WordPerfect® files module 72. The subgroup and join module 38 allows the user to select between a create subgroup module 74, a join subgroup module 76 and a join data module 78. The reclassify module 40 allows the user to select  
20 between a reclassify numeric variable module 80 and a reclassify character variable module 82. The functions associated with each of these modules is described below.

The user login module 20 controls access to protocol data. Before a user can log  
into the system, an administrator must set up the user identification and password, and  
25 specify the authorized protocol data.

The select protocol module 22 allows access to authorized protocols. Clinical research in the pharmaceutical industry is usually separated into drug, indication, and protocol levels, and the invention is designed to follow these conventions. An  
30 administrator must set up data access rights for each user for each specific drug,

indication, and protocol. Only authorized protocols can be accessed by the users. The users do not have the right to delete or modify any raw or analysis data provided by the administrator, but the user can create any in-process data sets or export data into other formats for further data manipulations.

5

The data explorer module 24 provides the interface to the various modules that allow the user to manipulate clinical data. Of these, the report and analysis module 30 allows access to the primary clinical data display modules.

10

Included under the report and analysis module 30 is the tables submodule 42, which is designed to create a variety of tables summarizing the data. Continuous variables can be summarized by a variety of statistics and can be grouped by class variables. The user can create 2-way cross-classification tables and can choose whether or not to include missing levels of a class variable in a table. Output from the tables submodule 42 can be customized by adding titles and/or footnotes. The user has the option to change the table font, and can choose to present the date and time of the output, the page number, and the page size (portrait or landscape). The tables can be saved and printed. There are three table styles to choose from labeled as first style 44, second style 46, and third style 48.

15

20

The first style table 44 is appropriate for summarizing continuous variables such as age, weight, and height. This style can be used to present up to 20 continuous variables with as many of the following statistics as the user wishes to present: number of non-missing, number of missing, range, sum, mean, variance, maximum, minimum, standard deviation, standard error of the mean, coefficient of variation, student's t for testing the null hypothesis that the mean is zero and the corresponding p-value; and corrected and uncorrected sums of squares. As an option the user can choose up to four grouping variables. This will create a separate table for each grouping variable combination.

25

30

The second style table 46 is appropriate for summarizing continuous variables such as age, weight, or height by classification variables such as race, sex, and treatment. This style can be used to present up to four continuous variables by two classification variables in a single table. A classification variable is required for second style and a  
5 single table is created, rather than a separate table for each combination of classification variables as in the first style table.

The third style table 48 is appropriate for creating a 2-way cross-classification table such as treatment by race. The table gives the frequency and percent of each  
10 combination of the classification variables. Percents can be presented as overall, row, or column percents. The user can choose up to two column classification variables and up to two row classification variables to be presented in a single table.

The graph submodule 58 is designed to create a variety of graphs summarizing  
15 the data. Output from the graph submodule 58 can be customized by adding a title and/or a footnote. Options such as title font are available in the menu bar. Available submodules included in the graph submodule 58 are mean 60, frequency 62, and plot 64.

The mean submodule 60 is designed to create horizontal and vertical bar charts,  
20 and 3D horizontal and vertical bar charts for one continuous variable, grouped by a classification variable. Subgrouping is also available. This submodule can be used to compare, for instance, the mean age between treatment groups, the mean age between genders, or the mean age between gender broken down by treatment groups. The output also includes the standard deviation of the response variable, as well as group frequency  
25 counts.

The frequency submodule 62 is designed to create horizontal and vertical bar charts, 3D horizontal and vertical bar charts, pie charts, and 3D pie charts. The response and grouping variables must be classification variables. Subgrouping is also available.  
30 For data sets that include the key variable, the user can choose between patient counts

or event counts. For data sets that do not include key variable, only event counts can be used. The user can chose to present the graph as frequency counts or as percents. The frequency submodule 62 can be used, for instance, to view the race distribution of subjects in a study, the gender distribution, or to view the race distribution among the treatment groups. One can compare the number of adverse events that occurred per treatment group, or the number of subjects who experienced an adverse event per treatment group.

The plot submodule 64 is designed to create line, scatter, or needle plots. The plot submodule 64 can be used, for example, to create a scatter plot of baseline versus final lab values to visually show a trend or to reveal outliers. The user can click on the outlier point to reveal such information as the patient number corresponding to the outlier, the treatment group that the patient is in, the patient's sex, age, and the x and y coordinates of the point. As another example the user can first create mean efficacy values by treatment and time using the statistics module 54. The plot submodule 64 can then be accessed from the statistics submodule 54 to create a line graph of efficacy values over time for each treatment group.

The listings submodule 50 is used to create data listings in a desirable format and layout. The variables listed in the output are limited and sorted by user specified variables. The user has the option to subset the data before creating a list. Output from the listings submodule 50 can be customized by adding titles and/or footnotes, or choosing from other available options.

The lab module 52 is used to view the laboratory data set. This module provides a gateway to explore the functionality in LAB<sup>®</sup> optionally implemented in some SAS<sup>®</sup> products.

The statistics submodule 54 is used to produce statistics for continuous data such as age, height, and weight. Available statistics are number of non-missing, number

of missing, range, sum, mean, variance, minimum, maximum, standard deviation, standard error of the mean, coefficient of variation, skewness, and kurtosis. Grouping is available in the statistics submodule 54. The graph submodule 58 can be accessed from the statistics submodule 54 so that the user can produce, for instance, a bar chart  
5 comparing mean age among treatment groups or mean change from baseline in efficacy variables among treatment groups. The user can also run the statistics submodule 54 to get mean efficacy values by treatment and time and then access the graph submodule 58 to create a line plot of mean efficacy values per treatment over time. Output from the statistics submodule 54 can be customized by adding titles and/or footnotes. The user  
10 can choose to present the date and time of the output, the page number, and the page size (portrait or landscape).

The insight submodule 56 is used to view the data set under the INSIGHT® module optionally implemented in some SAS products.

15

The browse data submodule 32 in the invention is designed to perform such functions as viewing, searching, sorting, saving, and printing data sets, and exporting data sets to an external file. The user can create new data sets from existing data sets with such options as select variables to keep, select variables to drop, and rename  
20 variable. The delete function allows the user to delete data sets he has created. Original data sets can not be deleted.

The find variable submodule 34 is used to identify the data sets that contain a particular variable. The find variable submodule 34 includes the ability to view data  
25 sets, to sort data sets, to save sorted data sets, to print data sets, and to export data sets to another file format. This module is linked to report & analysis submodule 30 and to the Browse Data Screen. The Browse Data Screen is similar to the browse data module 32 but lacks the Delete function.

30

The view files submodule 36 is used to view output files and tables created in the invention, to view other ASCII files, and to view WordPerfect files. Two submodules are available in the view files submodule 36: view files & tables 70, and view WordPerfect® files 72. From the view files & tables submodule 70 the user may view and print files and tables created in the invention and can choose options such as view font and page size (portrait or landscape). From the view WordPerfect files submodule 72 the user may invoke the software product WordPerfect.

The subgroup & join submodule 38 is organized into three groups: create subgroup 74, join & subgroup 76, and join data 78. These modules are linked to the report & analysis submodule 30 and the patient information module 26 so that functions available in these modules can be performed directly from the subgroup & join submodule 38. The create subgroup module 74 allows the user to create a subset from an existing data set and save the subset for later usage. For example, the user can create data sets containing all patients with adverse event equal to 'headache', containing only males, containing only those patients between 30 and 45 years of age, or a data set containing only males between the ages of 30 and 45. The join subgroup submodule 76 is designed to join a subgroup created using the create subgroup module 74 with another data set with the same key variable. This allows the user to create a data set including only the subjects which were identified in the create subgroup module 74. The join data module 78 allows the user to join two data sets together, by a key variable. The user can select all variables or selected variables to be kept in the new joined data set. This module also validates the data set that user selected.

The reclassify submodule 40 is used to create new variables from existing variables. The data set containing the new variables can be saved for later usage. The reclassify submodule 40 is linked to the report & analysis submodule 30 and the patient information module 26 so that functions available in these modules can be directly accessed from the reclassify submodule 40. Two submodules are available under the reclassify submodule 40: the reclassify numeric variable submodule 80 and the reclassify

character variable submodule 82. The reclassify numeric variable module 80 allows the user to reclassify numeric variables. As an example, the user can create a new classification variable 'newage' by grouping the numeric variable 'age' into levels such as <30, 30-45, and >45. The reclassify character variable module 82 allows the user to  
5 reclassify the variable according to the user's own grouping criteria. The user can select a data set and variables from the screen. All the existing values of the variable will be displayed for the user. The user can group different values of the variable to create a new variable.

10 The patient information module 26 allows the reviewer to browse the patient profiles. The Patient Information module also allows the user to view a subset of any data in the protocol, grouped by the subject identifier. It can also be invoked from other modules to view only the patient information in the current file that the user wishes to work with.

15 The adverse event module 28 allows the user to determine the frequency and percentage of patients with adverse events, grouped according to treatment, body system, and preferred term. It gets the information from an adverse event data set which was set up by the administrator, and displays the frequency and percentage values in an  
20 organizational chart format. The chart has three levels: treatment, body system, and preferred term. The module also provides the function to 'drill down' to a subset data set or to individual patient information.

In order to run the embodiment disclosed in the microfiche appendix, a stand  
25 alone or networked personal computer (PC) running Windows™ 3.1 or Windows™ for Workgroups 3.11 and MS-DOS® 6.x, or higher, would be sufficient. However, Windows 95™ or Windows NT™ platforms are recommended for optimum performance.

The *minimum* hardware requirements for the embodiment 100 disclosed in the appendix are: Intel<sup>®</sup> 486DX2 66 MHZ CPU 102, 256 KB cache, 32 MB RAM 104, a super VGA 15" 800X600 color monitor 106, and a Hewlett Packard (HP<sup>®</sup>) Laser printer 108 or InkJet printer. Minimum free hard disk space required depends on the size of the database and the location of the SAS software. A minimum of 100 to 150 MB of free hard disk space are recommended. As the possibility of running multiple applications increases, the recommended minimum free hard drive space will increase commensurately too. One representative hardware configuration that works well with the above-disclosed embodiment includes an Intel<sup>®</sup> Pentium 90 MHZ CPU, a 256 KB cache, 64 MB RAM, 150 MB free hard drive space, and a super VGA 17" color monitor.

As is obvious to those skilled in the art of computer software design, the above-disclosed invention could be readily adapted to operate on other computer platforms, including Unix<sup>®</sup> and MacIntosh<sup>®</sup> platforms.

One embodiment of the system was developed under SAS<sup>®</sup> System 6.11, and utilizes newly developed features like object oriented programming, data table object functions, and 'drag and drop' on-screen editing. The SAS<sup>®</sup> modules required for this application are SAS/BASE<sup>®</sup>, SAS/CORE<sup>®</sup>, SAS/AF<sup>®</sup>, SAS/FSP<sup>®</sup>, SAS/ACCESS<sup>®</sup>, SAS/STAT<sup>®</sup>, and SAS/GRAPH<sup>®</sup>. The invention provides the gateways to access SAS/INSIGHT<sup>®</sup> and SAS/LAB<sup>®</sup> modules. Competing programs that use other programming languages require conversion of SAS<sup>®</sup> data, which can cause errors. The invention eliminates this time-consuming and problematic step.

The SAS<sup>®</sup> data sets are defined for the most effective and efficient use. Although users can do extensive manipulation on the data sets, analysis data sets with the appropriate structure and sufficient information are provided. This prevents the user from spending unnecessary time manipulating data instead of reviewing data.



There are two basic data set layouts, which are referred to herein as "vertical" and "horizontal." In a vertical layout, a subject can have more than one associated record, whereas in a horizontal layout, each subject has only one record.

5 The following is an example of a vertical layout:

	TRT	SUBJECT	TIME	EFF VALUE
	tr2	3001	0	95
	tr2	3001	1	48
	tr2	3001	6	9
10	tr2	3001	24	22
	tr3	3002	0	77
	tr3	3002	1	53
	tr3	3002	6	30
	tr3	3002	24	61
15	tr1	3004	0	85
	tr1	3004	1	50
	tr1	3004	6	31
	tr1	3004	24	68

20 and the following is an example of the horizontal layout:

	TRT	SUBJECT	BASE	H1	H6	H24
	tr2	3001	95	48	9	22
	tr3	3002	77	53	30	61
25	tr1	3004	85	50	31	68

Both of these data sets contain the same information, but each subject in the second table contains only one record, whereas several records may be assigned to each subject in the first table.

Some data displays, such as graphs, require data in the vertical layout, while others will require the horizontal layout. When both layouts are provided, then the user can select the one that is required and quickly create the table or graph of interest.

5        Every data set contains variables that are frequently needed to create tables, graphs or lists. Demographic variables, such as sex, race and age, as well as the treatment code, are included in every data set. This saves the user time because he or she will not need to merge data sets together to get the needed variables into one data set. Likewise, efficacy data sets and lab data sets contain the change from baseline value  
10    at each time point.

The type of patient sample used to create graphs or tables can vary. For example, safety tables, such as adverse events, labs and vital signs may be created from the intent-to-treat safety sample. Efficacy tables may be created from the intent-to-treat  
15    efficacy sample. Indicator variables are included in every data set and can be used to extract the appropriate patient sample for the graph or table that the user wants to create.

The above described embodiments are given as illustrative examples only. It will  
20    be readily appreciated that many deviations may be made from the specific embodiments disclosed in this specification without departing from the invention. Accordingly, the scope of the invention is to be determined by the claims below rather than being limited to the specifically described embodiments above.

## CLAIMS

What is claimed is:

- 5 1. A method for processing clinical trial databases, comprising the steps of:
  - a. assembling data from at least one clinical trial database into a data file having a preselected format;
  - b. generating from the data file a table in a selected one of a first style, a second style or a third style; and
  - 10 c. generating a graph representing a selected subset of the data file.
2. The method of Claim 1, further comprising the steps of:
  - a. creating a subset from an existing data set; and
  - b. joining a subgroup created with another data set with having a same key
  - 15 variable.
3. An apparatus for processing clinical trial databases, comprising:
  - a. means for assembling data from at least one clinical trial database into a data file having a preselected format;
  - 20 b. means for generating from the data file a table in a selected one of a first style, a second style or a third style; and
  - c. means for generating a graph representing a selected subset of the data file.
- 25 4. The apparatus of Claim 3, further comprising:
  - a. means for creating a subset from an existing data set; and
  - b. means for joining a subgroup created with another data set with having a same key variable.

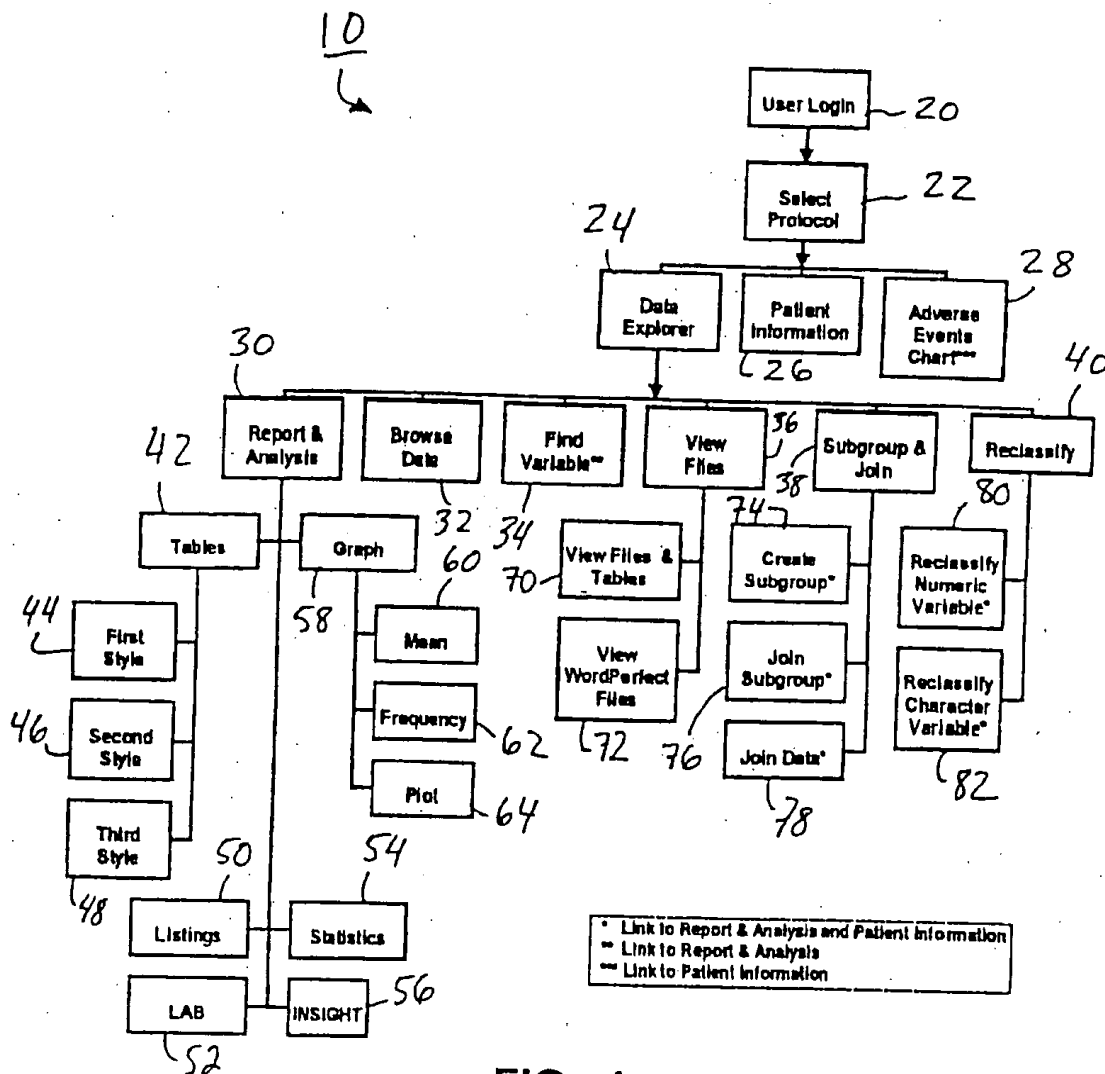
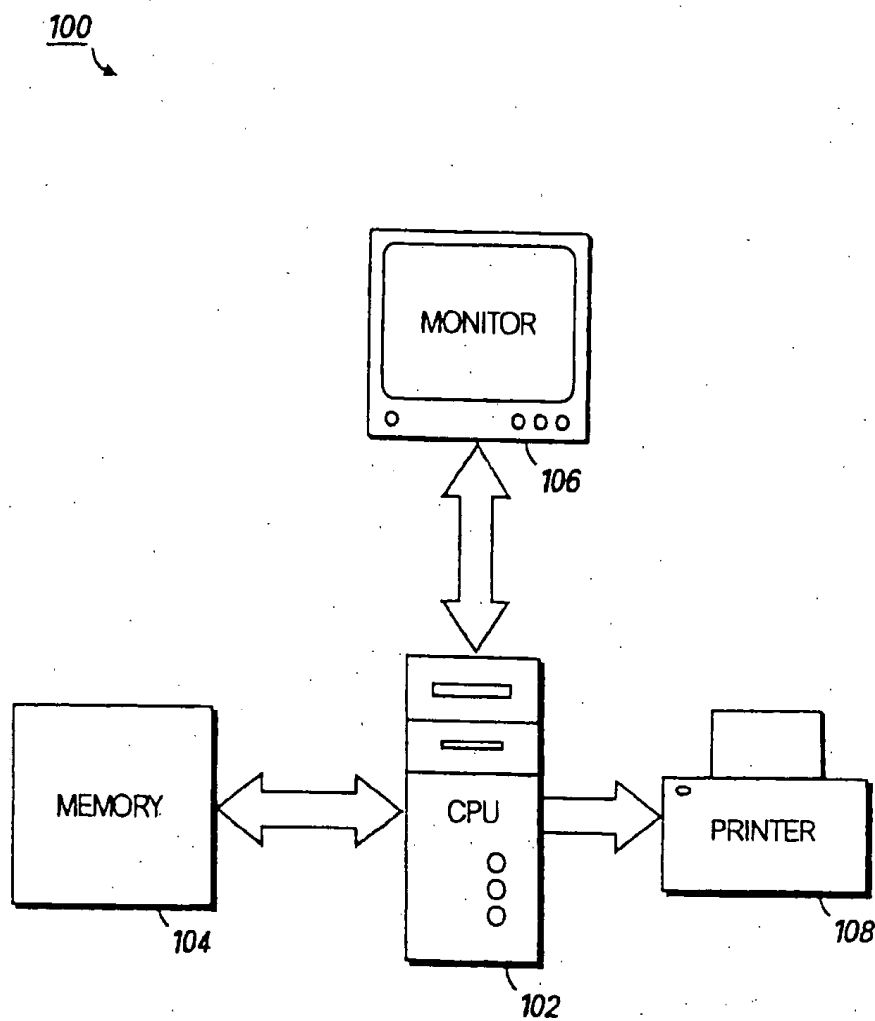


FIG. 1

**FIG.2**

SUBSTITUTE SHEET (RULE 26)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/16629

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06T 11/00; G06F 17/30

US CL : 345/440

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 345/440, 441, 352, 353

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	US 5,611,035 A (HALL) 11 March 1997.	NONE
X	US 5,461,708 A (KAHN) 24 October 1995, col.8, line 40 through col.12, line 29	1-4
A	US 5,345,544 A (TWASAKI ET AL) 06 September 1994.	NONE
A	US 5,515,488 A (HOPPE ET AL) 07 May 1996.	NONE
X	US 5,550,964 A (DAVOUST) 27 August 1996, col.7, line 59 through col.10, line 53.	1-4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* "A" document defining the general state of the art which is not considered to be of particular relevance	* "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* "B" earlier document published on or after the international filing date	* "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* "O" document referring to an oral disclosure, use, exhibition or other means	* "A" document member of the same patent family
* "P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

08 NOVEMBER 1997

Date of mailing of the international search report

26 JAN 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

OLIFF N. VO

Telephone No. (703) 305-9594